Applied Psychophysiology and Biofeedback (2021) 46:135–140
https://doi.org/10.1007/s10484-020-09502-y

Bayesian Data Analysis: A Fresh Approach to Power Issues and Null Hypothesis Interpretation

J. Peter Rosenfeld[1] · **Joseph M. Olson**[1]

Accepted: 31 December 2020 / Published online: 18 January 2021

Abstract
One of the first things one learns in a basic psychology or statistics course is that you cannot prove the null hypothesis that there is no difference between two conditions such as a patient group and a normal control group. This remains true. However now, thanks to ongoing progress by a special group of devoted methodologists, even when the result of an inferential test is $p > .05$, it is now possible to rigorously and quantitatively conclude that (a) the null hypothesis is actually unlikely, and (b) that the alternative hypothesis of an actual difference between treatment and control is more probable than the null. Alternatively, it is also possible to conclude quantitatively that the null hypothesis is much more likely than the alternative. Without Bayesian statistics, we couldn't say anything if a simple inferential analysis like a t-test yielded $p > .05$. The present, mostly non-quantitative article describes free resources and illustrative procedures for doing Bayesian analysis, with t-test and ANOVA examples.

Keywords Bayesian statistics · JASP · Probability ratios of alternative to null hypotheses

Background

One of the first things one learns in a basic psychology or statistics course is that you cannot prove the *null* hypothesis that there is no difference between two conditions such as a patient group and a normal control group. This remains true. However now, thanks to ongoing progress by a special group of devoted methodologists (called the "JASP team"), even when the result of an inferential test is $p > 0.05$—the test result was insignificant—it is now possible to *rigorously* and *quantitatively* conclude that (a) the null hypothesis is actually unlikely, and (b) the *alternative* hypothesis of an actual difference between treatment and control is more probable than the null. Alternatively, it is also possible to conclude *quantitatively* that the null hypothesis is much more likely than the alternative. Without Bayesian statistics, *we couldn't say anything* if a simple inferential analysis like a t-test yielded $p > 0.05$.

Moreover, again, with Bayesian statistics, it is also now possible to support the alternative hypothesis with far more confidence than the mere statement: "$p < 0.05$". One may now state *rigorously* that the alternative is (for example) 100 times more likely than the null! Before Bayesian inference, as promoted by the JASP team, even a $p < 0.05$ had to be taken with a grain of salt because, as my statistician colleagues emphasize in their elementary statistics courses, if one adds enough subjects to the contrasted cells of interest, eventually, one will arrive at the sacred $p < 0.05$. However, this may be unconvincing, and a statement about the greater relative likelihood of alternative than null is far more persuasive. The JASP team have provided methods that allow one to quantify the relative likelihoods of the alternative and null hypotheses. Indeed, a recent paper from the team (Keysers et al. 2020) handles the general argument about the need to discriminate a case of evidence for absence of an effect (i.e., in support of the null hypothesis) from a case of absence of evidence (i.e., a simple lack of support for the alternative hypothesis).

There have been many times when a rejected paper comes back with the comment that the null conclusion is unjustified because the submission was possibly underpowered, i.e.,

Supplementary information The online version of this article (https://doi.org/10.1007/s10484-020-09502-y) contains supplementary material, which is available to authorized users.

✉ J. Peter Rosenfeld
jp-rosenfeld@northwestern.edu

[1] Department of Psychology and Institute for Neuroscience, Northwestern University, Evanston, IL, USA

not enough participants (that are not always easy to amass, particularly in clinical neurofeedback studies) were studied. Indeed my own introduction to Bayesian analysis came in about 2014 when the editor of the prestigious journal, *Psychological Science,* reacted in exactly this way to one of our submissions (now published as Hu et al. 2015) with the demand that we apply Bayesian analysis –of which we had never before heard– to support our claim that cognitive suppression manipulations removed the effect of lying versus truth-telling on brain waves that we typically report. This demand prompted one of us (JPR) to register in 2015 for the superb workshop in Bayesian Statistics regularly offered *to all* at the annual meetings of the American Psychological Society. After that workshop, although we are no Bayesian experts, and never were statisticians, it is now easy for us to assimilate what we need to know in order to easily utilize this powerful new approach which we will try to share with you here.

The group of methodologists –the JASP team–noted above are those workers who have for the past decade or so been promoting a wholly different statistical approach—known in the 1930s, but neglected in psychological science until recently– based on Bayes' Theorem. They call themselves the JASP team in recognition of Bayesian statistics pioneer, Sir Harold Jeffreys, and "JASP" stands for "Jeffreys's Amazing Statistics Program." Bayes's theorem is named for Thomas Bayes (1701–1761), and there is an enormous literature about its history and derivation that may be easily found on the internet, just by using Google. This article will neither derive the theorem, nor attempt to demonstrate how the new statistical methodology follows from the theorem. Those chores are ably done in papers published by or cited in papers on the JASP team web page (http://pcl.missouri.edu/publications), such as Marsman and Wagenmakers (2017); Schönbrodt and Wagenmakers (2018); Wagenmakers et al. (2018a, 2018b). In the present little article, an attempt will be made to simply illustrate how easy it is to utilize and understand the resources that the JASP team has made available *at no charge!*

Indeed, if you simply browse to https://jasp-stats.org/, you will discover "JASP 0.14.1", a powerful software system that –after a *free* download– can do at no cost much of what SPSS, SYSTAT, or SAS can do (for charges of thousands of dollars)—plus a vast amount in addition. It will perform all manner of standard ANOVAs, ANCOVAs, Regressions, Correlations, and such, but will additionally do these analyses in a novel Bayesian manner that yields (in addition to the usual familiar values of F, t, r, p, effect sizes, etc.), novel terms called *Bayes Factors* (BFs)–that quantify the likelihood ratios of alternative and null hypotheses. Moreover, it easily inputs entire data tables in such familiar formats as .csv from EXCEL. It will also provide excellent tables and plots that can be copied and pasted directly into research submissions (as my colleagues and I have done in recent papers).

All this computing power is free and its graphical user interface is simple to learn. Moreover, JASP is an open source program to which users are constantly contributing. For example, there is an excellent toolbox for using machine learning (ML) algorithms, a collection of hot, new artificial intelligence approaches to developing diagnostic classifier functions that may be used, for example, to distinguish patient and control populations, based on various collected data from these groups. More familiar discriminant functions and linear regression approaches are the most basic form of such ML classifiers, that are far more powerful (see https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623).

JASP also provides a user forum, (https://forum.cogsci.nl/categories/jasp-bayesfactor) to which any one may pose questions about anything at all regarding use of the software, and typically receive useful answers usually within a day or two—again at no cost. Additionally, there is (1) a wonderful, free "How to use JASP" manual that one may find on line at https://jasp-stats.org/how-to-use-jasp/. (2) a wonderful free detailed "How to do it" guide manual at http://static.jasp-stats.org/Manuals/Bayesian_Guide_v0_12_2_1.pdf. They are chock full of information about how to use and interpret JASP. Finally, the latest version of JASP (0.14.1) also includes a preliminary module called "Learn Bayes" that introduces Bayesian statistics to newcomers.

Illustrations

T-Test

Let us give a few examples of how easy it is to get the BF—the likelihood ratio for Alternative (H1) with respect to Null (H0) Hypotheses–for a between-groups (independent samples) t-test. (Mathematically, this BF10 = [likelihood of obtaining the collected data given H1]/[likelihood of obtaining the collected data given H0]. Suppose you do such a test on two groups, each of size = 15, and you get a t-value of 2.2. This value at df = (15–2) = 13 yields $p < 0.05$ (*barely*, since the critical value of t is 2.106 for df = 13 at $p = 0.05$). To find the associated BF, one browses to the link provided above: http://pcl.missouri.edu/bayesfactor and the screen shown in Fig. 1 will appear:

Note in Fig. 1 that the "Bayes Factor Calculators" Tab in the blue taskbar at the top is selected. Note there is a set of calculator links shown, and the one we want is "Grouped or two-sample t-tests." If you click this link, the screen shown in Fig. 2 appears:

The default "Scale r" of 0.707 (explained in the JASP literature and links) is usually accepted. (If the user knows

a better value, s/he can enter it.) Now 15 is entered for n-values for both groups, and the t-value = 2.2 is entered, then one clicks "submit". The result that appears on the screen is this:

Two-Sample Design

Input is:
N1 = 15
N2 = 15
t = 2.2
Scale r = 0.707

Bayes factor in favor of the **alternative**:

Scaled JZS Bayes Factor = 1.993054
Scaled-Information Bayes Factor = 2.758125

That BF ("JZS Bayes Factor") of about 2 ("1.993"), despite the t-value being significant at $p < 0.05$, is *not* strong evidence for the alternative hypothesis, although it does mean that the data set collected under the alternative hypothesis is about twice as likely as data set collected under the null hypothesis. What exactly does the value imply? There is a table shown in Table 1 from Schönbrodt and Wagenmakers (2018) that tells us.

In this table, as usual "H0" means the null hypothesis and "H1" means the alternative, and we can see that our BF value of about 2 (favoring alternative) is merely anecdotal (though worth mentioning as such), despite $p < 0.05$. If you wish, try entering larger n or t values. In the case of $t = 2.2$, you will find that doubling participant numbers to 30 per group will not much improve the BF, but a t value of 2.5 with the same 15 plus 15 participants yields a $BF = 3.22$, which, as shown in Table 1, is moderate evidence for the alternative.

Let us now consider the case of a non-significant ($p > 0.05$) result. Suppose one has 70 per group and the t-value is not significant at $t = 1.1$. The t-value required for significance at $p < 0.05$ is about 1.99, so $t = 1.1$ is *ns*. Of course p values never make it possible to say how likely a true null hypothesis is, so one has run 140 subjects to no clear end. However, if one enters the group sizes and t into the BF link given above, one gets this output on the screen:

N1 = 70
N2 = 70
t = 1.1
Scale r = 0.707
Bayes factor in favor of the null:
Scaled JZS Bayes Factor = 3.173656.

This is moderate, *quantitative* evidence for the null that is computed to be greater than three times as likely as the alternative. Table 1 gives fractional values for the null hypothesis. These are reciprocals of the values for alternatives. Thus 1/3.173656 = 0.315, which is about 3/10. This too is said to be moderate evidence for the null in Table 1 above.) The Bayes Factor provides moderate evidence for the null without an increase in power, because the Bayes Factor does *not* provide an uninterpretable probability statement about the null hypothesis (i.e., $p > 0.05$), as the t-test does, but rather provides a ratio of probabilities of null to alternative. Because the BF supporting Null [BF(01)] gives the ratio of the predictive performance of the null to that of the alternative, it is possible to provide quantitative evidence that the null model is more likely than the alternative model, indicating that the analysis was sufficiently powered to demonstrate the lack of an effect. In a case where an analysis is truly underpowered, however, the Bayes Factor will be indeterminate (e.g., BF of less than 2, meaning alternative is about as likely as null), preventing the researcher from drawing any firm conclusions from insufficient evidence (see Schönbrodt and Wagenmakers 2018 for an excellent discussion of Bayesian power analyses). It is noted that the BF in support of the alternative (H1) is called BF (10) and the BF in support of the null (H0) is called BF (01). These BFs are reciprocals of each other; i.e., BF (10) = 1/BF (01) and BF (01) = 1/BF (10). (We would finally note that all these analyses illustrated above may be conducted also in JASP 0.14.1, whose "Summary Statistics" module provides additional further information).

Table 1 Bayes factor interpretation, from Schönbrodt and Wagenmakers (2018)

Bayes factor	Evidence category
> 100	Extreme evidence for $\mathcal{H}_1$
30 - 100	Very strong evidence for $\mathcal{H}_1$
10 - 30	Strong evidence for $\mathcal{H}_1$
3 - 10	Moderate evidence for $\mathcal{H}_1$
1 - 3	Anecdotal evidence for $\mathcal{H}_1$
1	No evidence
1/3 - 1	Anecdotal evidence for $\mathcal{H}_0$
1/10 - 1/3	Moderate evidence for $\mathcal{H}_0$
1/30 - 1/10	Strong evidence for $\mathcal{H}_0$
1/100 - 1/30	Very strong evidence for $\mathcal{H}_0$
< 1/100	Extreme evidence for $\mathcal{H}_0$

Bayesian Analysis of Variance (ANOVA)

What about a more complicated design than one requiring a simple t-test, say a 2 × 2 factorial? Let us say the two factors are A and B, where the two levels of A are drug treatment vs. biofeedback, and the two levels of B are patients vs. normals. If one does a standard ANOVA (as found in SPSS), it is easy to obtain the BFs for both A and B main effects: Since t = the square root of F, all one needs to do is take the square root of the F values for A and B—yielding t– and use the BF calculator link supplied above as before.

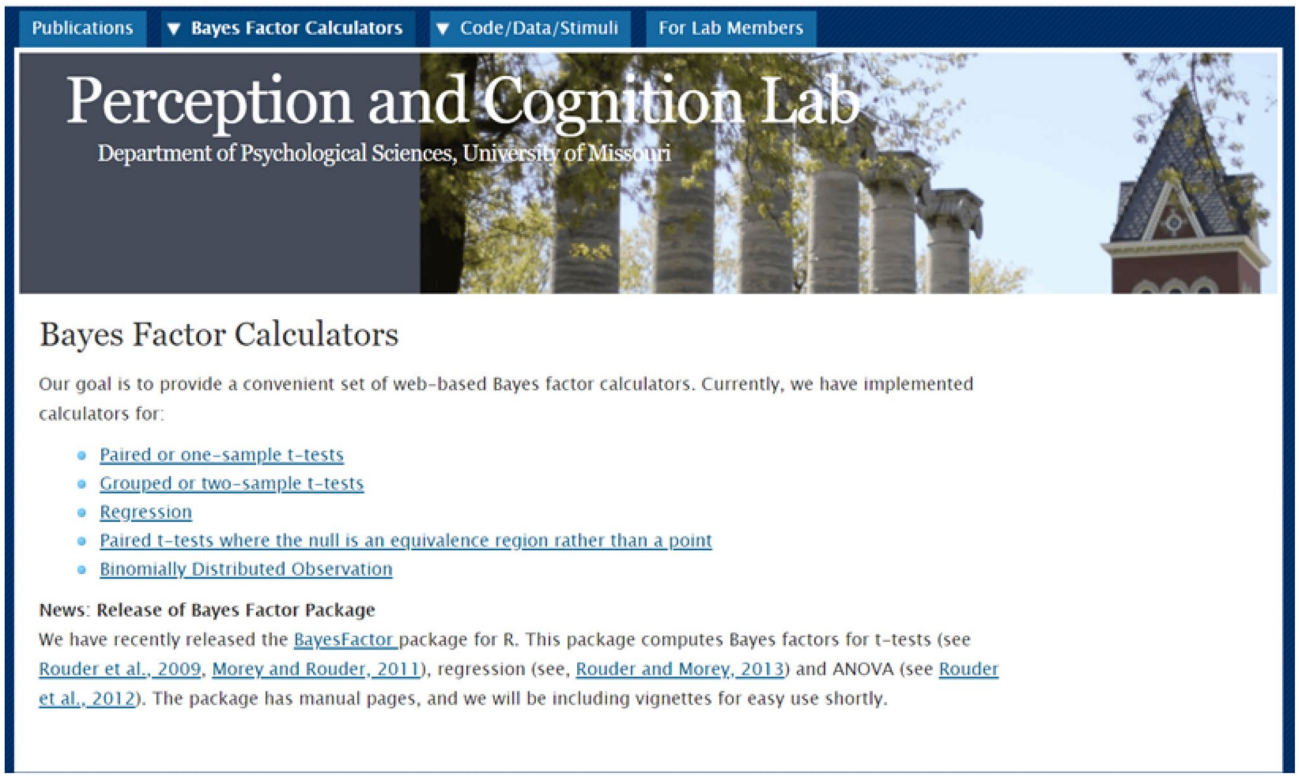

Fig. 1 Bayes factor calculators (This wonderful link was provided by Jeffrey Rouder, Professor of Cognitive Sciences at the University of California, Irvine)

However, there is no similarly straightforward way to find the BF either in support of the interaction or a BF providing support against the interaction. To get this information, one must do a Bayesian ANOVA as supplied in the free JASP software package described above. It is not hard at all after one has done it once or twice. We will not go through it step by step here, with multiple screenshots, because it is a bit more involved. It involves inputting the data set into the JASP worksheet, naming the independent variables (factors), and calling for the Bayesian ANOVA that is appropriate, e.g., totally independent (aka *completely between groups*) in the example given above. (Of course, repeated measures, and mixed ANOVAs are also available.) One can do both standard and Bayesian ANOVAs in JASP, and one should do both.

Because the Bayesian ANOVA is somewhat more complex than the t-test, the output of this test is also a bit more complex—yet still easily interpretable. The Bayesian ANOVA outputs Bayes Factors (and other probabilities) for each model of the data one may conceptualize, depending upon which factors have significant effects on the data. For example in the 2-factor model suggested above (factor A; Biofeedback vs. Drugs and factor B; Patients vs. Controls), there are various possible hypotheses or models that may explain the obtained data: (1) a model suggesting that the data are influenced only by factor A, (2) a model suggesting that the data are influenced only by factor B, (3) a model suggesting that the data are influenced by both factors A and B independently (4) a model suggesting that the data are influenced by factors A and B independently *plus* the interaction of A and B (A*B). After telling JASP to do a Bayesian ANOVA on the submitted data, and calling for BFs (BF10s) in support of alternative hypotheses compared to the null, part of the critical output from JASP is in the form shown in Table 2.

The numerical values here and below are hypothetical and illustrative. The "Models" in the first column are the same as the models given just above involving factors A, B, A + B, and (A + B + A*B). The probability of the models prior to data collection, P (M), are unknown so each is equally weighted at 0.2 (This is called an "uninformed prior"). The probability of each model, given the data, has been calculated and represented in the "P (M/data)" column. P (M/data) means the conditional probability of the model given the newly collected data.) These likelihoods are used to calculate the Bayes Factors given in the "BF10" column (highlighted in yellow). To illustrate the simple calculation of the BF10, we will contrast the model suggesting that A is the sole relevant factor affecting the data with the Null model.

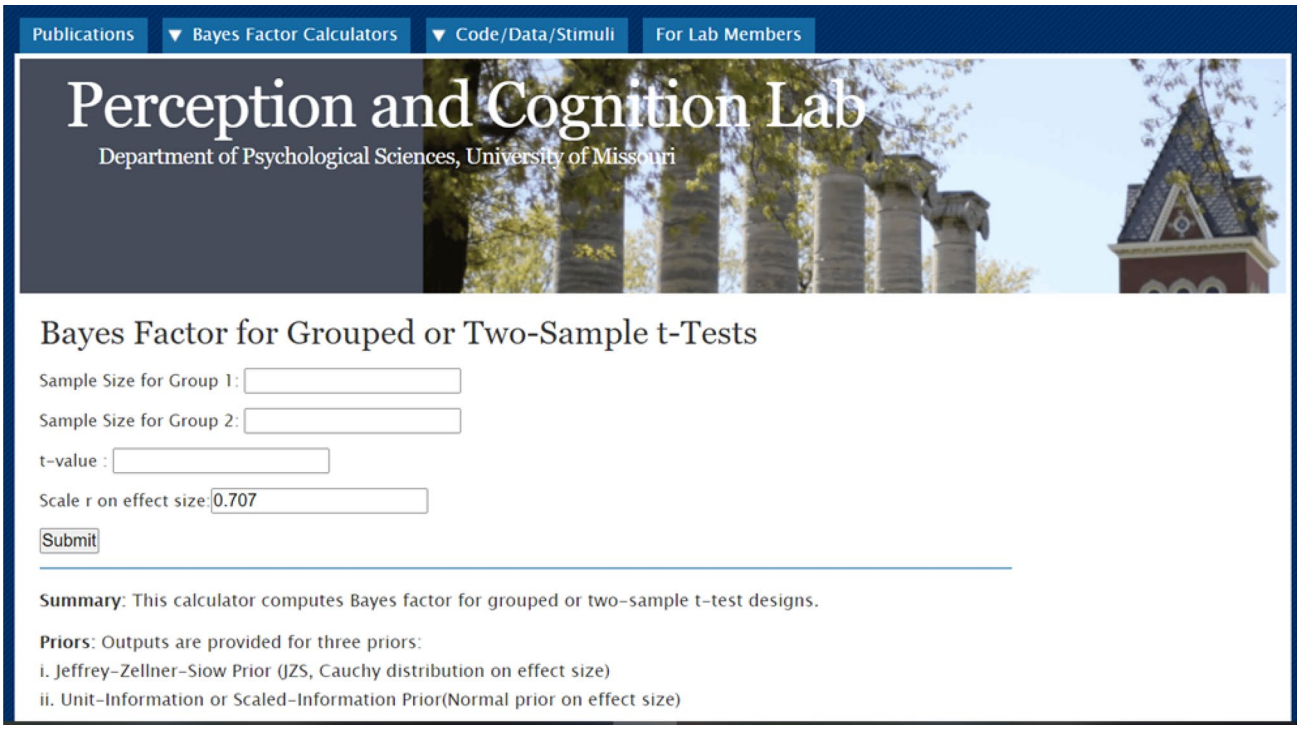

Fig. 2 Computing the Bayes factor for a between-groups t-test

Table 2 Bayesian ANOVA output 1

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	7.640e −13	3.056e −12	1.000	
B	0.200	2.487e −13	9.949e −13	0.326	0.836
A	0.200	0.642	7.163	8.399e +11	0.677
A + B	0.200	0.291	1.645	3.814e +11	1.188
A + B + A*B	0.200	0.067	0.287	8.760e +10	1.445

Note. All models include subject

Table 2 above shows that the probability of the null model has been calculated to be "7.640e -13," (which means 7.64 with the decimal point shifted 13 places left), that is, extremely unlikely given the observed data. The probability of model A, on the other hand, has been calculated to be relatively high, at "0.642." Given these probabilities, it is possible for us to calculate the BF10 for model A, which represents how probable the data are under model A compared to how probable the data are under the null model. Dividing 0.642 by 7.640e -13 will demonstrate that model A is 8.399e + 11 (8.399 times ten raised to the 11th power) times more likely than the null model – decisive evidence for model A. (The supplement explains that we have here made certain simplifying assumptions that allow us to compute the Bayes Factor in this way.) This is why the Bayes Factor is often described as a "likelihood ratio" given that the Bayes Factor represents the likelihood of the data under one model (e.g., model A) as compared to the likelihood of the data under another (in this example, the null model). Additionally, the BFs supporting the Null hypotheses for each factor (BF01) may be found by simply taking the reciprocals of the BF10 values (or by initially requesting them in submitting the data set to JASP).

What should be immediately apparent is that there is no separate BF10 output for the interaction factor, A*B. However, JASP also outputs a *special and perhaps novel kind of BF called BF* (incl) for this interaction in another table shown in Table 3.

This Bayes factor (= 0.230) is the likelihood ratio of a model containing the interaction (e.g., A + B + A*B)

Table 3 Bayesian ANOVA, output 2

Analysis of Effects ▼

Effects	P(incl)	P(incl|data)	BF_{incl}
A	.400	.933	9.2e +11
B	.400	.291	.454
A*B	.200	.067	.230

Note. Compares models that contain the effect to equivalent models stripped of the effect.

with respect to a model excluding it (e.g., A + B). That excluding model is called BF (excl). These new BFs are also interpreted using Table 1 given above, but substituting BF(incl) for BF (10) and BF (excl) for BF(01), in this example, with BF (incl) = 0.23 providing no support for the inclusive model. However its reciprocal is *BF* (excl), and that value = 1/0.23 = 4.35, which states that the exclusive model is 4.35 times more likely than the inclusive model. This is moderate evidence that the factors A and B are not interacting.

Similar to BF10, the BF (incl) is the ratio of the likelihood of one model compared to the likelihood of another. For example, to calculate the BF (incl) for the interaction (= 0.230), one simply takes the likelihood of the model with the interaction (model A + B + A*B = 0.067) in the Table 2 above and divides it by the likelihood of the identical model, but lacking the interaction (model A + B = 0.291), which yields BF (incl) = 0.230, as in Table 3 also output from JASP. (Here again, please see the supplement).

The two fractional BF (incl) values (for B and, as already noted, the AxB interaction) in Table 3 are < 1, so provide no support for the significance of those effects. The BF in support of the significant effect for factor A is overwhelming; 9.2e + 11, which is mathematical notation for 9.2 followed by 11 zeros, which, being > 100, is said to be "extreme" evidence in the table of BFs from Schönbrodt and Wagenmakers (2018) given above as Table 1.

Conclusions

The JASP team provides all this material at no charge because, as they explicitly state in their papers and web sites, based on compelling logical evidence, they believe that Bayesian analysis is the correct data-analytic approach, and they would like it if all the research community saw things the same way. We present authors are clearly converts.

References

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: Evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science, 26*(7), 1098–1106.

Keysers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience, 23,* 788–799.

Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology, 14*(5), 545–555.

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review, 25*(1), 128–142.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., & Matzke, D. (2018a). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin and Review, 25*(1), 35–57.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., & Meerhoff, F. (2018b). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review, 25*(1), 58–76.